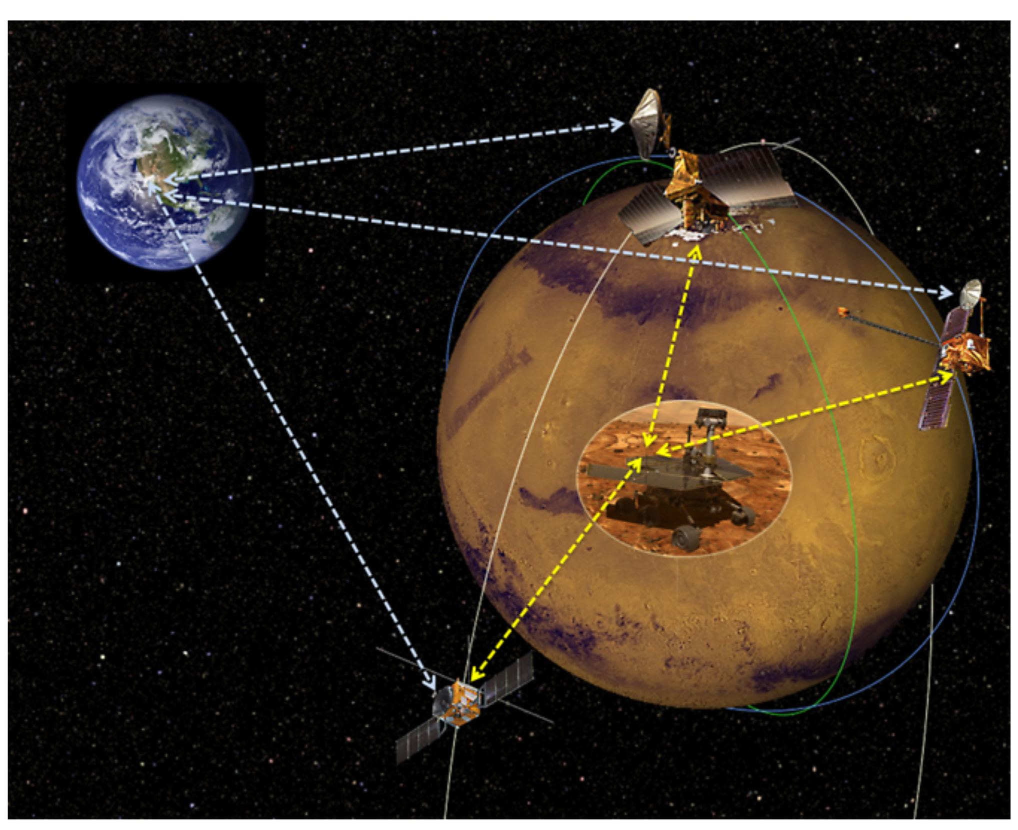
National Aeronautics and Space Administration

Automated Data Accountability for the Mars Science Laboratory

Brian Kahovec (393K),
Ryan Alimo (393K), Dariush Divsalar (332B)

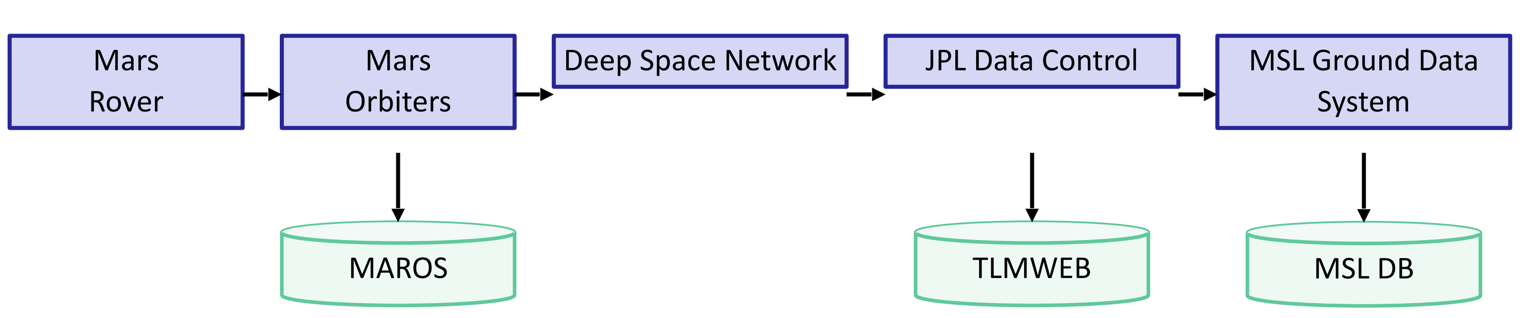
Introduction

Data Accountability is the process of ensuring that all data sent from a spacecraft is received and processed successfully on the ground and also identifying where in the pipeline data becomes missing if it not received. The Mars Science Laboratory (MSL) currently relies on Ground Data Systems Analysts (GDSA) to determine whether or not all data has been accounted for. When data is missing, it can take several hours to determine the root cause of the issue. There have been previous attempts to automate the data accountability process, but they are unreliable for operational use. This paper presents machine learning based approaches to automate and optimize the detection of volume loss from the downlink process of telemetry data from the Mars Curiosity Rover.



MSL Downlink Process

During the downlink process, data is transferred from the rover to one of the Mars orbiters. The orbiter then sends the data to one of the Deep Space Network Stations. The station sends the data to the Jet Propulsion Laboratory, where it is received by Data Control and stored in the Telemetry Data System (TDS). Finally, the MSL Ground Data System (GDS) software processes the data from the TDS and stores it in the MSL Database.



Approach

Data Collection

A data pipeline accumulates information about each downlink at various locations in the MSL Downlink Process. We gather data from three data sources: MAROS, which contains metadata from one of the orbiters; TLMWeb, which contains metadata from JPL Data Control, and the MSL Database, which contains metadata after the downlink has been processed by GDS software.

Signal Processing

A signal processor combines the raw metadata from these three separate sources and computes the relevant features. Expert GDSAs evaluated the computed features for each transmission to label each downlink as complete or incomplete. Our labelled dataset consists of approximately 9000 downlinks.

Machine Learning

With our well-labelled dataset, we trained both supervised learning models to classify each downlink as Complete or Incomplete. We also applied unsupervised learning techniques to identify anomalies in the data. These anomalies are equivalent to Incomplete passes labelled by the supervised learning algorithms. Our dataset is imbalanced; only about 10% of passes are Incomplete or anomalous, so detecting these can be difficult but is important for the GDSA team.

Hyperparameter Optimization

To further increase the accuracy of our models, we applied hyperparameter optimization while training. The Variational Autoencoder (VAE) had the highest recall of incomplete passes. Since our dataset is imbalanced, identifying these anomalies is more difficult and this recall is a better metric than overall accuracy. We chose to retrain the VAE with various hyperparameter optimization algorithms.

Accuracy of Trained Models

Machine Learning Algorithm	Recall of Incomplete Passes	Overall Accuracy
Adversarial Autoencoder	62%	67%
Variational Autoencoder	**80%**	**77%**
Linear Regression	38%	92%
Support Vector Machine	68%	85%
Gaussian Naïve Bayes	27%	97%
Deep Neural Network	50%	91%

Results of the trained machine learning models without hyperparameter optimization

Optimization Algorithm	Recall of Incomplete Passes	Overall Accuracy
Random Search	91%	83%
Tree Parzen Estimator	96%	87%
Hyper NOMAD	92%	85%
Delta-DOGS	94%	79%
Delta-MADS	**97%**	**88%**

Results of different optimization algorithms applied while training the Variational Autoencoder

Infusion and Explainability

The most accurate model was delivered to the MSL GDSA team. The model was infused into their software and is used to perform their daily operations. When a pass is labelled as Incomplete, the GDSAs need to know why. To answer this question, we determine which feature is the most anomalous by computing the error of each feature. Then an error message tells the GDSA which feature is anomalous and and the value of the feature. This error message provides the GDSA enough information to respond to the issue.

Sol 2696					
46960	TGO_MSL_2020_066_03	Complete		350.273
36960	MRO_MSL_2020_066_04	Complete		316.189
46961	TGO_MSL_2020_067_02	Complete		185.804
36961	MRO_MSL_2020_067_01	Incomplete	The data volume difference between the orbiter and TDS is 0.13 MB	397.564
36961	MRO_MSL_2020_067_01	Complete		397.654
Sol 2697				
46970	TGO_MSL_2020_067_03	Complete		391.194
36970	MRO_MSL_2020_067_02	Complete		467.631

The MSL GDSA Report Summary Dashboard displays the status of each downlink as computed by the trained model. An error message explains where the data was lost.

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov